An effective Bayesian model for lithofacies estimation using geophysical data

Jinsong Chen and Yoram Rubin

Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA

Received 15 August 2002; revised 12 February 2003; accepted 5 March 2003; published 7 May 2003.

[1] A Bayesian model coupled with a fuzzy neural network (BFNN) is developed to enhance the use of geophysical data in lithofacies estimation. Prior estimates are inferred from borehole lithofacies measurements using indicator kriging, and posterior estimates are obtained by updating the prior using geophysical data. The novelty of this study lies in the use of the fuzzy neural network for the inference of the likelihood function. This allows spatial correlation of lithofacies as well as nonlinear cross correlation between lithofacies and geophysical attributes to be incorporated into lithofacies estimation. The effectiveness of BFNN is demonstrated using synthetic data emulating measurements at the Lawrence Livermore National Laboratory (LLNL) Site. *INDEX TERMS:* 1829 Hydrology: Groundwater hydrology; *KEYWORDS:* lithofacies, geophysical data, neural networks, Bayesian, statistical model, fuzzy computing

Citation: Chen, J., and Y. Rubin, An effective Bayesian model for lithofacies estimation using geophysical data, *Water Resour. Res.*, *39*(5), 1118, doi:10.1029/2002WR001666, 2003.

1. Introduction

[2] Heterogeneity of lithofacies has an important effect on the determination of hydrogeological and geochemical parameters in flow and contaminant transport [*Dagan*, 1989]. Modeling this heterogeneity often requires the use of multiple sources of information, especially noninvasive and cost-effective geophysical data [*Keys*, 1997; *Hubbard and Rubin*, 2000; *Kowalsky et al.*, 2001]. Geophysical methods such as seismic reflection or refraction surveying delineate shallow subsurface structure by detecting interfaces between different lithofacies. These methods are often used to map large-scale subsurface structures, such as bedrock and channels, where the contrast of geophysical properties between adjacent structures is large. To provide detailed information about the spatial distribution of lithofacies, however, methods such as borehole logging and cross-hole tomographic surveying may be beneficial. Such methods map lithofacies from their corresponding geophysical properties, which are measured either directly at the boreholes or indirectly from geophysical signals that traverse between the boreholes.

[3] Several problems exist in the development of models for incorporating geophysical data as well as geological and hydrogeological information into lithofacies estimation. The first problem is how to relate geophysical attributes to lithofacies under field conditions. Typically, these relations are nonunique and site-specific. Another problem is how to quantify uncertainty associated with lithofacies estimation due to ambiguity of geophysical interpretation, measurement errors, and the scale disparity between different types of geophysical measurements [*Rubin et al.*, 1992]. Deterministic models, such as graphical methods [*Doveton*, 1986], neural networks methods [*Rogers et al.*, 1992], and fuzzy neural networks methods [*Chang et al.*, 1997], have limitations in terms of these criteria.

[4] Geostatistical methods, especially indicator-based models, have been found to be effective in overcoming the aforementioned problems. For example, indicator kriging and sequential indicator simulation methods can be used to estimate or simulate lithofacies at unsampled locations based on spatial correlation of lithofacies. These methods can reproduce the spatial patterns of lithofacies revealed from measured data without considering depositional processes [*Koltermann and Gorelick*, 1996]. The methods, however, lack a means of incorporating potentially useful geophysical information into the estimation procedure. Consequently, methods such as indicator cokriging [*Rosenbaum et al.*, 1997; *Deutsch and Journel*, 1998], simulated annealing [*Aarts and Korst*, 1989; *Deutsch and Cockerham*, 1994], and Bayesian methods [*Rubin et al.*, 1992; *Copty and Rubin*, 1995; *Chen et al.*, 2001] have been developed to combine geophysical and direct sampled data.

[5] Bayesian models are flexible in combining multiple sources of information and have been successfully used for many years to incorporate geophysical data into lithofacies estimation. Bayes' theorem provides a general framework for data assimilation and allows various types of information to be integrated in a hierarchical manner [*Box and Tiao*, 1973; *Bernardo and Smith*, 1994]. Using carefully built Bayesian models, surface and crosshole geophysical data as well as borehole lithofacies and geophysical logs can be jointly used in lithofacies estimation. The parameters of those Bayesian models can be identified using optimization models, such as the maximizing a posteriori probability density function (MAP) method [*Lortzer and Berkhout*, 1992], or sampling-based models, such as the Markov chain Monte Carlo (MCMC) method [*Gilks et al.*, 1998; *Bosch*, 1999; *Bosch et al.*, 2001].

[6] However, one problem in using Bayesian models is in the inference of the likelihood functions. Since relations

0043-1397/03/2002WR001666$09.00

between lithofacies and geophysical attributes are often complex, it is difficult to make assumptions about the form of the likelihood functions in advance. Instead, in this study we develop a nonparametric model (a fuzzy neural network) to derive the likelihood functions directly from a given data set, with very few assumptions made about the form of the likelihood function.

[7] The remainder of this paper is organized as follows. Sections 2 and 3 describe the development of a general Bayesian model and a fuzzy neural network for estimating the likelihood function, respectively. Sections 4 and 5 allow for evaluation of the effectiveness of the model in combining geophysical data by using two sets of synthetic data that are based on field measurements at the Lawrence Livemore National Laboratory (LLNL) Site. Conclusions are given in section 6.

2. Bayesian Model

2.1. Bayesian Framework

[8] The proposed Bayesian model is developed based on typical situations of environmental site characterization, such as those at the Lawrence Livemore National Laboratory (LLNL) Site [*Ezzedine et al.*, 1999]. The goal of the study is to estimate lithofacies at location $\mathbf{x}$, given lithofacies measurements at locations $\mathbf{x}_i$, $i \in A$, and given various types of geophysical data $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $\cdots$, $g_t(\mathbf{x})$ at location $\mathbf{x}$, where $A = \{1, 2, \cdots, n\}$ is an index set, n is the total number of lithofacies measurements, and t is the total number of geophysical attributes. The geophysical data, made dimensionless after normalization, can be obtained from crosshole tomographic surveys or estimated from intensive borehole geophysical measurements using kriging. Let $Z(\mathbf{x})$ be a discrete random variable taking the value of 1, 2, $\cdots$, or up to q, where q is the total number of lithofacies at a site. Let $z(\mathbf{x}_i)$ be the lithofacies at $\mathbf{x}_i$, an integer value between 1 and q. The conditional probability to observe lithofacies of type k at location $\mathbf{x}$ can be determined by applying the following Bayesian formula:

$$\begin{aligned} P(Z(\mathbf{x}) = k | g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_t(\mathbf{x}), z(\mathbf{x}_i), i \in A) \\ = [Cf(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_t(\mathbf{x}) \mid Z(\mathbf{x}) = k, z(\mathbf{x}_i), i \in A)] \\ P(Z(\mathbf{x}) = k | z(\mathbf{x}_i), i \in A), \end{aligned} \tag{1}$$

where f is a joint conditional probability density function, referred to as a likelihood function, and C is a normalizing coefficient included so that the summation of the conditional probability for different types of lithofacies equals one.

[9] We can simplify the Bayesian model given in Equation (1) by dropping $z(\mathbf{x}_i)$, $i \in A$, from the likelihood function under the Markov assumption [*Almeida and Journel*, 1994]. Let $I_k(\mathbf{x})$ be an indicator random variable, defined by

$$I_k(\mathbf{x}) = \begin{cases} 1 & \text{if } Z(\mathbf{x}) = k \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

and therefore $\{Z(\mathbf{x}) = k\}$ is equivalent to $\{I_k(\mathbf{x}) = 1\}$. We replace $\{Z(\mathbf{x}) = k\}$ with $\{I_k(\mathbf{x}) = 1\}$ in Equation (1), and let $P_{prior}(I_k(\mathbf{x}) = 1) = P(I_k(\mathbf{x}) = 1 | z(\mathbf{x}_i), i \in A)$, referred to as prior probability, and $P_{post}(I_k(\mathbf{x}) = 1) = P(I_k(\mathbf{x}) = 1 | g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_t(\mathbf{x}), z(\mathbf{x}_i), i \in A)$, referred to as posterior probability. This leads to

$$\begin{aligned} P_{post}(I_k(\mathbf{x}) = 1) = [Cf(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_t(\mathbf{x}) | I_k(\mathbf{x}) = 1)] \\ \cdot P_{prior}(I_k(\mathbf{x}) = 1). \end{aligned} \tag{3}$$

[10] In Equation (3), we present a Bayesian approach for obtaining the probability of the kth lithofacies occurring at location $\mathbf{x}$, using lithofacies measurements and geophysical data. The lithofacies measurements were used to infer the prior probability, whereas the geophysical data were used to update the prior through the likelihood function. We can use the obtained posterior probability to estimate lithofacies at each location, as we did later in the two case studies, and the uncertainty of the estimation is determined by $Var(I_k(\mathbf{x})) = P_{post}(I_k(\mathbf{x}) = 1) \cdot [1 - P_{post}(I_k(\mathbf{x}) = 1)]$, where $Var(\cdot)$ is the variance of the indicator variable $I_k(\mathbf{x})$ [*Stone*, 1995]. We can also use the posterior probability to generate lithofacies random fields, using the sequential indicator simulation methods [*Deutsch and Journel*, 1998]. Note that the probability of the kth lithofacies occurring at location $\mathbf{x}$ itself can be considered as a random variable defined on [0, 1], and estimation of the probability is also subject to uncertainty. This is not our focus in the paper, and the reader interested in the topic is referred to *Stone* [1995].

2.2. Prior Probability

[11] The prior probability in Equation (3) can be estimated from lithofacies measurements, $z(\mathbf{x}_i)$, $i \in A$, using indicator kriging [*Rosenbaum et al.*, 1997; *Deutsch and Journel*, 1998]. Let p_k be the unconditional probability of the kth type of lithofacies occurring at location $\mathbf{x}$, which can be estimated from borehole lithofacies measurements. The prior probability thus is given by the simple kriging estimator:

$$P_{prior}(I_k(\mathbf{x}) = 1) = p_k + \sum_{i \in A} \lambda_i(\mathbf{x})(I_k(\mathbf{x}_i) - p_k). \tag{4}$$

The coefficient $\lambda_i(\mathbf{x})$, $i \in A$, are obtained by solving the following equation:

$$\sum_{i \in A} \lambda_i(\mathbf{x}) C_I(\mathbf{x}_i, \mathbf{x}_j) = C_I(\mathbf{x}, \mathbf{x}_j), \quad j \in A, \tag{5}$$

where $C_I(\mathbf{x}_i, \mathbf{x}_j)$ and $C_I(\mathbf{x}, \mathbf{x}_j)$ represent the covariance of the indicator variables at locations $\mathbf{x}_i$ and $\mathbf{x}_j$, and at locations $\mathbf{x}$ and $\mathbf{x}_j$, respectively.

[12] In this study, we use only lithofacies measurements at boreholes to determine the prior probability. Other types of information such as geologic interpretation and expert experience can also be incorporated into estimation of the prior [*Rubin*, 2003]. As the distances between locations $\mathbf{x}$ and $\mathbf{x}_i$, $i \in A$, become large, the conditional probability will be close to the unconditional probability p_k, and lithofacies measurements at boreholes will have little influence on the estimation.

2.3. Likelihood Function

[13] The likelihood function in Equation (3) is the link between lithofacies and geophysical attributes, and can be inferred using collocated lithofacies and geophysical data.

The inference, however, is challenging because correlations between lithofacies and geophysical attributes are often nonlinear, and covariance-based likelihood functions [*Kitanidis*, 1986; *Copty and Rubin*, 1995] may not be applicable.

[14] As an alternative to the covariance-based likelihood functions, we propose a nonparametric approach. Let $h(\mathbf{x}, \boldsymbol{\theta})$ be a function derived from a fuzzy neural network with several decision rules, where $\boldsymbol{\theta}$ is a vector of parameters. Let $l(\mathbf{x})$ be the logarithm of the likelihood function (log-likelihood) given by

$$l(\mathbf{x}) = \log(Cf(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_t(\mathbf{x}) | I_k(\mathbf{x}) = 1)). \quad (6)$$

We can approximate $l(\mathbf{x})$ with h, which can be identified from a training data set $(g_1(\mathbf{x}_i), g_2(\mathbf{x}_i), \cdots, g_t(\mathbf{x}_i), l(\mathbf{x}_i))$, $i \in$ A, by minimizing:

$$\sum_{i \in A} [h(\mathbf{x}_i, \theta) - l(\mathbf{x}_i)]^2. \quad (7)$$

The log-likelihood $l(\mathbf{x})$ at an arbitrary location $\mathbf{x}$ in Equation (7) is unknown, but it can be estimated at measurement locations $\mathbf{x}_i$, $i \in A$, as described in the next section.

[15] The method for determining likelihood functions in this study is different from the maximum likelihood method used by *Kitanidis and Vomvoris* [1983], *Rubin and Dagan* [1988, 1989], *Kitanidis* [1997], and *Rubin* [2003]. They assumed a Gaussian probability model with the mean and covariance of the distribution as parameters and identified the parameters by searching for those values that maximize the probability of observing the measurements. In our approach, however, no assumption is made with regard to the form of the likelihood function or the posterior probability. Instead, we seek a nonparametric function to approximate the true log-likelihood by minimizing the sum of the squared residuals between the true log-likelihood and the approximated values at measurement locations. We use the obtained nonparametric function to replace the true log-likelihood in lithofacies estimation.

3. Fuzzy Neural Network

[16] A fuzzy neural network is a model based on fuzzy sets and fuzzy reasoning [*Takagi and Sugeno*, 1985]. A fuzzy set is a collection of elements with a continuum of grades of membership [*Rojas*, 1996]. It provides an intuitive way to describe imprecise information or ambiguous categorization. For example, a cluster plot of $g_1(\mathbf{x})$ versus $g_2(\mathbf{x})$ in an ideal situation can be classified as a few well-defined crisp clusters. Most real-life situations, however, are defined by overlapped fuzzy clusters as shown in Figure 1. Probabilistic tools can assign probability for one interpretation or another, but such approaches preclude interpretation such as "close to B but with some features of A in it". This is where the methods developed by fuzzy logic can come in handy.

[17] Fuzzy reasoning is similar to human thinking and implemented using a set of inference rules. Humans do not always comprehend numbers well but understand a collection of data that can be assembled or associated meaningfully by consequence of their similarity, resemblance, or operational cohesion [*Hirota and Pedrycz*, 1999]. Following this idea, a fuzzy neural network first classifies original data into several clusters and then applies the results to fuzzy rules for inferences. This model looks simple, but it is very effective and robust in modeling complex relations [*Zadeh*, 1994].

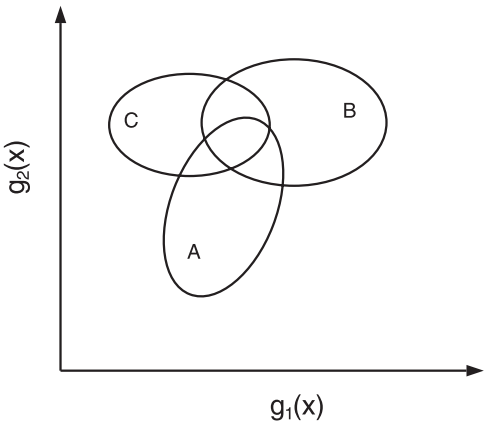

Figure 1. Examples of fuzzy clusters.

[18] In our application, we do not employ the advantage of fuzzy logic in describing imprecise information; rather, we benefit from it as a tool for modeling nonlinear correlations. We believe that our approach opens the door for using fuzzy logic to deal with large varieties of information and complex relations among them. In the remainder of this section, we will first introduce the structure of the fuzzy neural network and then outline the learning algorithm for parameter identification. Our focus is on model implementation rather than intuitive interpretation of parameters.

3.1. Structure

[19] The fuzzy neural network used in this study is similar to the one given by *Takagi and Sugeno* [1985] and *Jang* [1993]. To illustrate the network, let us consider two types of geophysical attributes with two possible decision rules. Let $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ be geophysical attributes at location $\mathbf{x}$. Using the notations employed by *Zadeh* [1965], the rules are defined as follows: rule 1, If $g_1(\mathbf{x})$ is A_1 and $g_2(\mathbf{x})$ is B_1, then the consequence is o_1; rule 2, if $g_1(\mathbf{x})$ is A_2 and $g_2(\mathbf{x})$ is B_2, then the consequence is o_2. Each inference rule includes two parts, the premise and the consequence. A_i and B_i ($i = 1, 2$) in the premises of those rules are linguistic labels, such as small and large, or fuzzy sets, defined by membership functions $\mu_{A_i}[g_1(\mathbf{x})]$ and $\mu_{B_i}[g_2(\mathbf{x})]$:

$$\mu_{Ai}[g_1(\mathbf{x})] = \exp\left(-\left(\frac{g_1(\mathbf{x}) - c_{i1}}{\sigma_{i1}}\right)^2\right),$$
$$\mu_{Bi}[g_2(\mathbf{x})] = \exp\left(-\left(\frac{g_2(\mathbf{x}) - c_{i2}}{\sigma_{i2}}\right)^2\right), \quad (8)$$

where c_{i1} and c_{i2} are the centers of fuzzy sets A_i and B_i, and σ_{i1} and σ_{i2} are parameters measuring the spread of data around their corresponding centers, referred to as the bases of the fuzzy sets. The variables o_1 and o_2 in the consequences of the fuzzy rules are determined from training data sets during the learning process, which can be constants or linear functions of input $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ [*Takagi and Sugeno*, 1985]. The choice of the consequences depends on the complexity of applications. For those applications involving highly nonlinear functions, we shall use linear functions of input as the consequences. For simplicity, we use constants

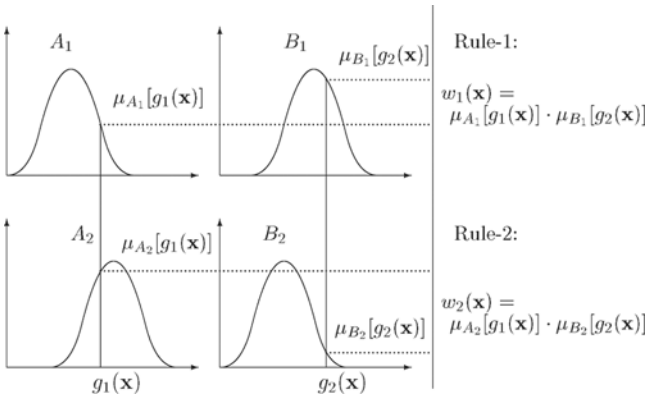

Figure 2. Premises of the fuzzy inference rules.

as the consequences in this study, but the developed methods are applicable to the consequences that are linear functions of input.

[20] Figure 2 shows the premises of the two inference rules in the fuzzy neural network. The weights of rule-1 and rule-2 at location $\mathbf{x}$ are determined using geophysical data $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$. We first compute the grade of membership $\mu_{A_i}[g_1(\mathbf{x})]$ of $g_1(\mathbf{x})$ in fuzzy sets A_i and the grade of membership $\mu_{B_i}[g_2(\mathbf{x})]$ of $g_2(\mathbf{x})$ in fuzzy set B_i. We then determine the weights as follows:

$$w_i(\mathbf{x}) = \mu_{Ai}[g_1(\mathbf{x})] \cdot \mu_{Bi}[g_2(\mathbf{x})], \qquad i = 1, 2. \tag{9}$$

Although other methods may also be used to compute the weights, the current method is computationally most efficient [*Takagi and Sugeno*, 1985]. The output of the network is a weighted linear combination of each consequence as given by

$$h(x, \theta) = \frac{w_1(\mathbf{x})o_1 + w_2(\mathbf{x})o_2}{w_1(\mathbf{x}) + w_2(\mathbf{x})}, \tag{10}$$

where $\boldsymbol{\theta}$ is a vector of parameters containing c_{i1}, c_{i2}, σ_{i1}, σ_{i2}, and o_i (i = 1, 2). This system is very efficient in fitting nonlinear functions. As shown by *Jang* [1993] and *Rojas* [1996], it can be used to approximate any continuous function defined on a bounded domain. One reason for the efficiency is that the weights are nonlinear functions of $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$, which are based on the patterns inherent in training data sets.

[21] Figure 3 shows the structure of the fuzzy neural network in which the computational process from one layer to another is illustrated. The input to each node in layer 1 is $g_1(\mathbf{x})$ or $g_2(\mathbf{x})$, and the output is the corresponding grades of membership. The outputs of the nodes in layer 2 and 3 are the weights $w_i(\mathbf{x})$ and relative weights $\bar{w}_i(\mathbf{x}) = w_i(\mathbf{x})/(w_1(\mathbf{x}) + w_2(\mathbf{x}))$, respectively. The input to each node in layer 4 is the relative weight $\bar{w}_i(\mathbf{x})$ and the consequence o_i, and the output of each node is the product of its corresponding input. The node in the last layer is the summation of all incoming signals to the node as given by $h(\mathbf{x}, \boldsymbol{\theta}) = \bar{w}_1(\mathbf{x})o_1 + \bar{w}_2(\mathbf{x})o_2$, which is equivalent to Equation (10).

3.2. Model Calibration

[22] The illustrative example in the last section contains only two inference rules. In real situations, however, before applying the fuzzy neural network to the problem of lithofacies estimation, we need to determine the number of inference rules required and identify from the data the parameters associated with those rules.

3.2.1. Structure Identification

[23] The intention of structure identification is to determine the number of inference rules and to provide the initial values of the center and base of each fuzzy set. We will use the fuzzy C mean (FCM) cluster analysis method [*Bezdek*, 1981] to identify possible patterns in the input data space. Let $(g_1(\mathbf{x}_j), g_2(\mathbf{x}_j))$, $j \in A$, be an input data set, and suppose we want to divide this data set into m clusters. Using the K means algorithm [*Hartigan and Wong*, 1979] or model-based clustering methods [*Banfield and Raftery*, 1993], we can divide the data set into m crisp subsets, in which each element in the data set can only belong to one of the subsets. Such methods are limited because they do not allow an element to partially belong to more than one subset.

[24] The fuzzy C mean algorithm divides the input data set into m fuzzy clusters, each of which is a fuzzy set in the sense that the boundaries between sets are poorly defined and possibly overlap. Any data point may partially belong to several fuzzy clusters with different grades of membership. Let $\mathbf{c}_i = (\mathbf{c}_{i1}, \mathbf{c}_{i2})^T$ be the vector of centers of the ith fuzzy cluster, $\mathbf{y}_j = (g_1(\mathbf{x}_j), g_2(\mathbf{x}_j))^T$ be the jth input data point, and u_{ij} be the grade of membership of $\mathbf{y}_j$ in the ith fuzzy cluster, where $i = 1, 2, \cdots, m$ and $j \in A$. The algorithm for the clustering can be outlined as follows.

[25] 1. Randomly assign the initial grades of membership $u_{ij} \in [0, 1]$ under the constraint $\sum_{i=1}^{m} u_{ij} = 1$.

[26] 2. Compute the center of each fuzzy cluster using

$$\mathbf{c}_i = \frac{\sum_{j \in A} u_{ij}^{\phi} \mathbf{y}_j}{\sum_{j \in A} u_{ij}^{\phi}}, \qquad i = 1, 2, \cdots, m, \tag{11}$$

where $\phi > 1$ is the fuzziness index and the default value is 2 [*Bezdek*, 1981].

[27] 3. Update the grades of membership using the newly obtained centers $\mathbf{c}_i$, i = 1, 2, $\cdots$, m. Let $d_{ij} = \| \mathbf{c}_i - \mathbf{y}_j \|$ and $d_{kj} = \| \mathbf{c}_k - \mathbf{y}_j \|$. If $d_{ij} = 0$, let $u_{ij} = 1$ and $u_{kj} = 0$ for $k \neq i$; otherwise, let

$$u_{ij} = \frac{\left(1/d_{ij}\right)^{\frac{2}{\phi-1}}}{\sum_{k=1}^{m} \left(1/d_{kj}\right)^{\frac{2}{\phi-1}}}. \tag{12}$$

[28] 4. Compare the newly updated grades of membership with the old ones. If they are close to each other, stop the iteration; otherwise, go back to step 2.

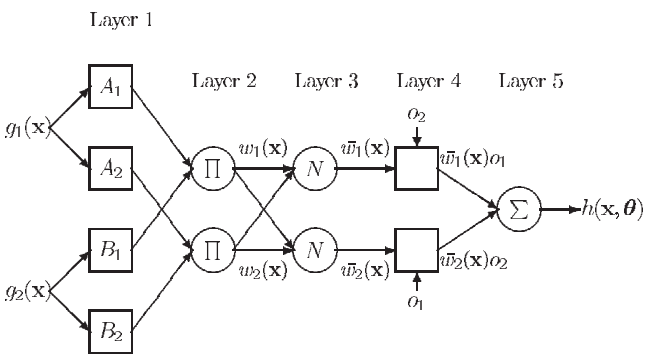

Figure 3. Structure of the fuzzy neural network.

[29] The number of fuzzy rules is determined by considering both compactness and separation of fuzzy clusters [*Xie and Beni*, 1991]. The compactness of the clusters is defined as the weighted squared distance as given by

$$J(\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_m) = \sum_{i=1}^{m} \sum_{j \in A} u_{ij}^{\phi} \parallel \mathbf{c}_i - \mathbf{y}_j \parallel^2 . \quad (13)$$

The separation of the clusters is defined as the minimum distance between the centers of each pair of clusters as follows

$$d_{min} = \min_{i \neq j} \parallel \mathbf{c}_i - \mathbf{c}_j \parallel, i,j = 1, 2, \cdots, m. \quad (14)$$

Both compactness and separation decrease with the increasing of the number of fuzzy clusters. For a good cluster analysis, data points within clusters are expected to be compact, whereas the cluster centers are expected to be well separated. Consequently, the number of clusters can be determined by minimizing the function

$$S = \frac{J(\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_m)/n}{d_{min}^2}, \quad (15)$$

where n is the total number of data points. The minimization of S is straightforward, as we know the number of clusters is greater than 1 but less than the total number of data points n. We first calculate S for $m = 2, 3, \cdots, n - 1$, and then select the value of m that gives the minimal value of S as the maximum number of clusters.

3.2.2. Parameter Identification

[30] After determining the number of fuzzy rules, which is equal to the number of fuzzy clusters, we need to identify the parameters associated with each of the rules. For example, in the case of two inference rules as shown in Figure 2, we need to estimate centers and bases of all fuzzy sets, c_{i1}, c_{i2}, σ_{i1}, σ_{i2}, and the consequences of each rule o_i, where $i = 1, 2$. We started from the initial values of c_{i1}, c_{i2}, σ_{i1}, σ_{i2}, where c_{i1} and c_{i2} are obtained from Equation (11) and σ_{i1} and σ_{i2} are given by the sample deviations of the clusters.

[31] The training data in the example consist of geophysical data ($g_1(\mathbf{x}_i)$, $g_2(\mathbf{x}_i)$), the input to the fuzzy neural network system, and the corresponding log likelihood $l(\mathbf{x}_i)$, the output of the system. The log likelihood value is not a measurement but can be estimated based on a simple postulate. Using Equations (3) and (6) and replacing $\mathbf{x}$ with $\mathbf{x}_i$, we obtain

$$l(\mathbf{x}_i) = \log\left[P_{post}(I_k(\mathbf{x}_i) = 1)\right] - \log\left[P_{prior}(I_k(\mathbf{x}_i) = 1)\right]. \quad (16)$$

The prior probability $P_{prior}(I_k(\mathbf{x}_i) = 1)$ can be estimated from lithofacies measurements at other boreholes using indicator kriging, and the posterior probability $P_{post}(I_k(\mathbf{x}_i) = 1)$ is assigned (as postulated) according to the lithofacies measurement at location $\mathbf{x}_i$. Special care should be taken when the prior and posterior probabilities are close to zero since log transformation of zero is invalid. We can use the following rules to avoid zero values of prior and posterior probabilities.

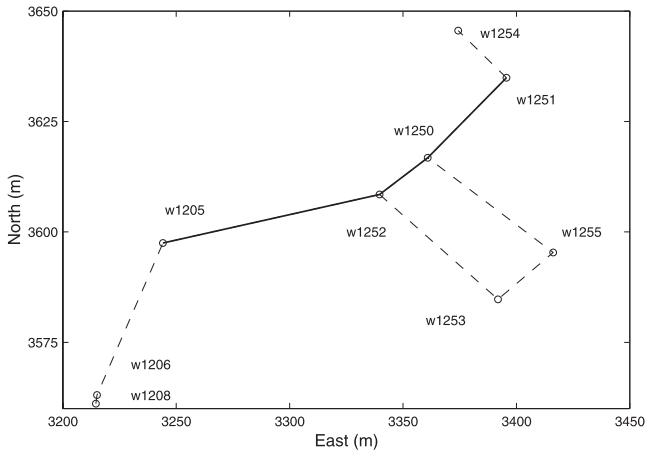

Figure 4. Locations of the well bores on the area near treatment facility D (TFD) at the Lawrence Livemore National Laboratory Site in California. The circles denote well bores, and the solid lines denote the profiles along which the synthetic data are generated.

1. Let $P_{prior}(I_k(\mathbf{x}_i) = 1) = \varepsilon$, if $P_{prior}(I_k(\mathbf{x}_i) = 1) < \varepsilon$.
2. Let $P_{post}(I_k(\mathbf{x}_i) = 1) = 1 - \varepsilon$, if $z(\mathbf{x}_i) = k$; let $P_{post}(I_k(\mathbf{x}_i) = 1) = \varepsilon$, otherwise.

In this model, we employed ε between 0.01 and 0.05. Although the value of ε may seem to be arbitrary, the estimated results are not sensitive to it. Details of the algorithm for determining parameters is given in Appendix A.

4. Case Study 1

[32] The first case study demonstrates the effectiveness of BFNN in combining borehole and cross-hole data for lithofacies estimation, using a synthetic data set generated based on field measurements at the Lawrence Livemore National Laboratory (LLNL) Site in California. The hydrogeology of the site is very complex, but a considerable amount of geological, geophysical, hydrogeological, and geochemical data are available [*Ezzedine et al.*, 1999]. These provide us a unique opportunity to study the use of geophysical data for shallow subsurface characterization.

4.1. Synthetic Data

[33] Our focus is on the area near treatment facility D (TFD) at the LLNL site shown in Figure 4. Of all the geophysical logs collected at the site, gamma ray and electrical resistivity logs were found to be most informative for lithofacies identification. Table 1 summarizes the geostatistical properties of the lithofacies, gamma ray, and electrical resistivity obtained from borehole measurements within hydro-stratigraphic unit 2 (HSU2). The gamma ray measurements were converted into shaliness to remove inconsistencies associated with data acquisition [*Doveton*, 1986], using the method described by *Ezzedine et al.* [1999]. Lithofacies and gamma ray shaliness have large correlation lengths, while electrical resistivity is spatially uncorrelated.

[34] Synthetic data were generated along a profile between wells 1205 and 1251 as indicated by a solid line in Figure 4, using the parameters listed in Table 1. We first generated a two-dimensional lithofacies field by condition-

Table 1. Statistical Properties of Lithofacies and Geophysical Data

	Lithofacies		Gamma Ray	Resistivity
	Sand	Silt	(Shaliness)	(Ωm)
Proportion	0.52	0.48	NA	NA
Horizontal range(m)	30.0	30.0	25.0	NA
Vertical range(m)	1.50	1.50	2.50	NA
Nugget	0.0	0.0	0.011	σ_R^2
Sill	0.25	0.25	0.040	σ_R^2
Variogram model	exponential	exponential	Gaussian	NA

ing to lithofacies measurements at the boreholes using the sequential indicator simulation method [*Rubin*, 2003]. We then generated a two-dimensional gamma ray shaliness field by conditioning to both the previously generated lithofacies field and the gamma ray shaliness available at the boreholes using the sequential Gaussian simulation method [*Deutsch and Journel*, 1998]. Finally, by conditioning to the previously generated lithofacies and gamma ray shaliness fields, we obtained a two-dimensional electrical resistivity field along the same profile.

[35] We divided the synthetic data set into two subsets, one for training and the other for testing. The training subset was created by randomly selecting eight boreholes as shown in Figure 5. The testing subset is the remaining synthetic data. Figure 6 shows a comparison of the normalized indicator experimental and theoretical variograms. The consistency between the two variograms confirms that the generated lithofacies field is a realization of the random field with the spatial structure given in Table 1. Figure 7 shows a scatterplot of gamma ray shaliness versus electrical resistivity based on data at the eight boreholes. We notice that the cross correlation between gamma ray shaliness and electrical resistivity is nonlinear and nonunique.

4.2. Approach

[36] The advantages of using BFNN for lithofacies estimation are investigated through comparison with several other models commonly used for site characterization. Each model is first calibrated using the data at the eight boreholes, and the calibrated model then is used to estimate lithofacies at each testing location. We compare the estimated results with their corresponding true values to eval-

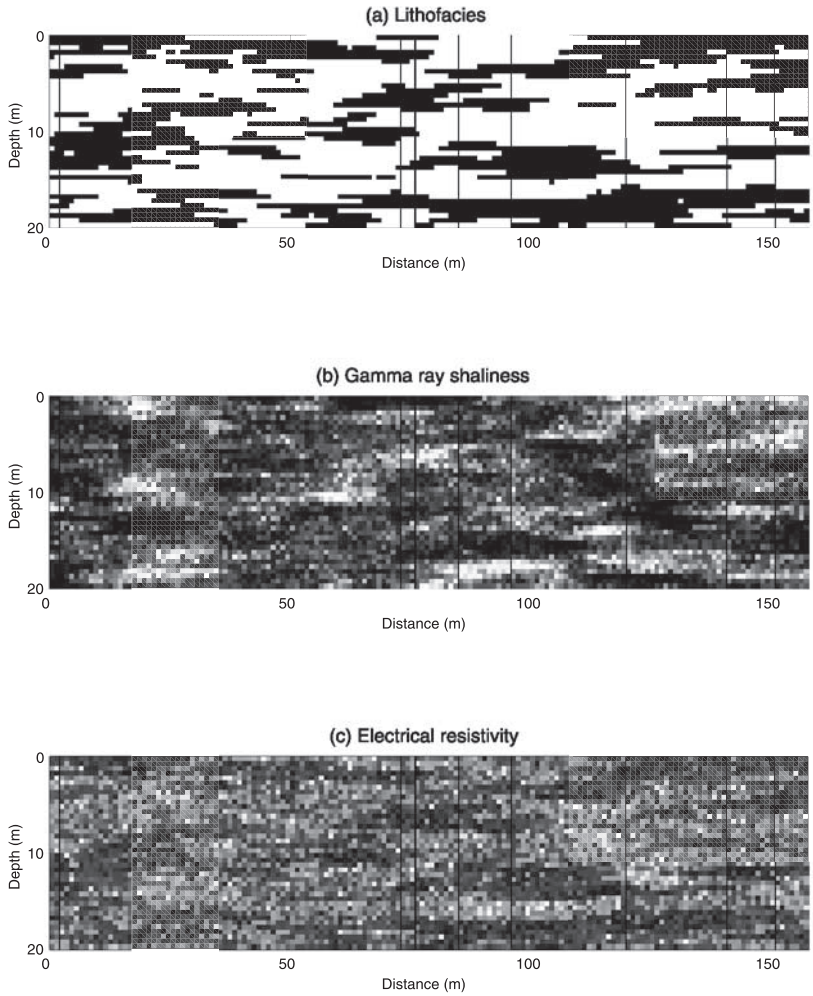

Figure 5. Simulated (a) lithofacies field with silt (black) and sand (white), (b) gamma ray shaliness field with values between 0 (black) and 1 (white), and (c) electrical resistivity field with values between 5Ωm (black) and 30Ωm (white), based on the field measurements at the LLNL site.

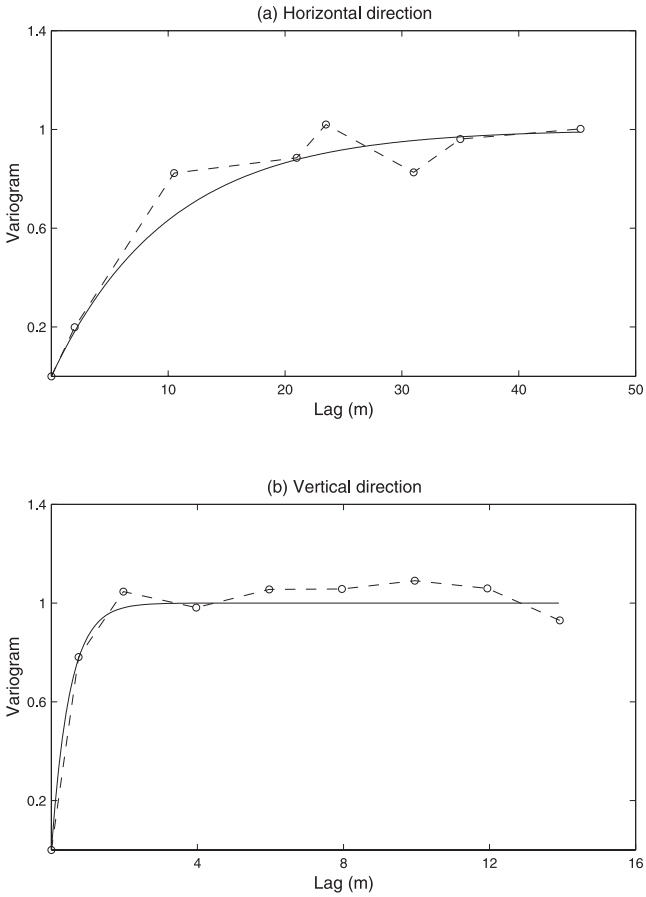

Figure 6. Variograms (a) along the horizontal direction and (b) along the vertical direction. The dashed lines with circles are the experimental variograms based on the data sampled at the eight boreholes, and the solid lines are the theoretical ones used for generating the synthetic data.

uate the efficiency of each model. To further explore the effects of the spatial correlation, we count the total number of misclassification as a function of the minimum distances between the testing and measurement locations.

[37] The models used for comparison include indicator kriging, indicator cokriging, and the fuzzy neural network (FNN) that does not consider spatial correlation of lithofacies [*Chang et al.*, 1997]. The indicator kriging is described in section 2.2, and the indicator cokriging is similar to the method given by *Almeida and Journel* [1994], where only collocated geophysical data are used. Since the indicator variable and the shaliness vary between 0 and 1, we normalize the electrical resistivity data to a range between 0 and 1 by first subtracting their minimum value and then dividing by their range. The detailed model is given in Appendix B.

4.3. Results

[38] Figure 8 shows the relative changes in misclassification with increase in the minimum distances between the testing and measurement locations using the different approaches. For FNN model, the percentage of misclassification does not depend on the minimum distances because lithofacies at any location is estimated only from the corresponding collocated geophysical data. Consequently, although the method is as efficient as BFNN model when the testing locations are far from the measurement locations, it does not perform as well as the testing locations get closer to the measurement locations.

[39] In the case of indicator kriging, misclassification also increases with the minimum distances as shown in Figure 8. As the distances increase, the lithofacies measurements have limited direct influence on lithofacies estimation. As a result, misclassification tends to a constant value determined by the unconditional probability of each type of lithofacies. This is reasonable as indicator kriging is based on spatial correlation of lithofacies, and it cannot consider information at a large distance. However, BFNN clearly benefits more from geophysical data at large distances from boreholes.

[40] Figure 8 suggests that indicator cokriging performs similar to BFNN model in terms of the percentages of misclassification, but this may not be always the case. Indicator cokriging is a covariance-based predictor that depends on correlation and cross correlation coefficients. In this case study, there are only two types of lithofacies, and the nonlinearity of the cross correlations between the geophysical attributes and lithofacies is somewhat reduced. The advantages of using BFNN model compared to the indicator cokriging are therefore not obvious. This aspect will be investigated in the next section.

5. Case Study 2

[41] The second case study is designed to evaluate the effects of nonlinear cross correlations between lithofacies and geophysical attributes on the performances of BFNN

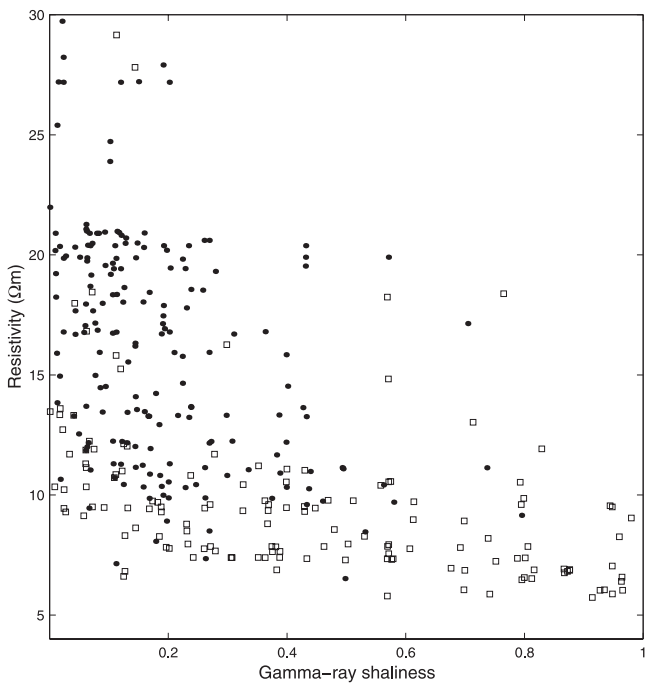

Figure 7. Scatterplot of gamma ray shaliness versus electrical resistivity based on the data sampled at the eight boreholes. The solid dots denote sand, and the squares denote silt.

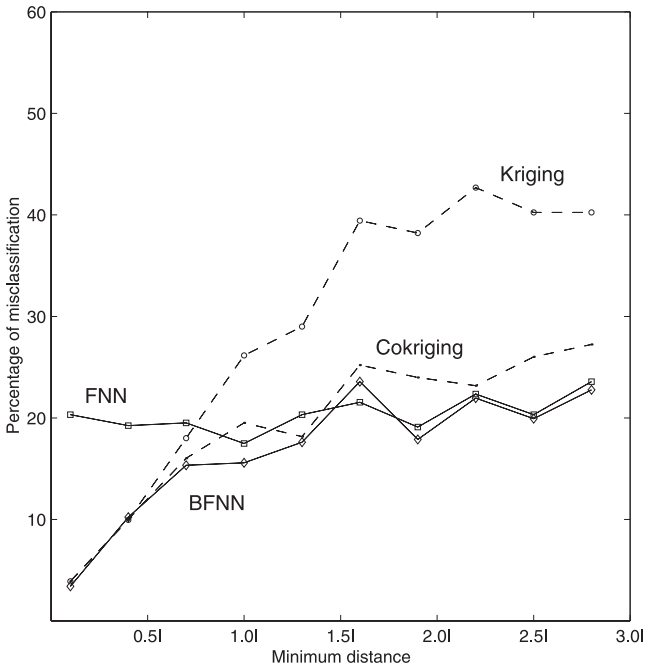

Figure 8. Comparison of misclassification using the synthetic data set, where I = 10 m is the integral length of sand along the horizontal direction.

and indicator cokriging models. Since the nonlinearity of the cross correlations usually increases with the number of lithofacies types, we will conduct our analysis on synthetic data with two, three, and four types of lithofacies.

5.1. Synthetic Data

[42] Synthetic data for the case study include three lithofacies fields shown in Figure 9 and geophysical data generated along five boreholes shown as solid lines. Each field is of dimension 120 m × 40 m and generated using the indicator sequential simulation method. The indicator variograms for all types of lithofacies were modeled as an exponential model with an integral scale of 10m in the horizontal direction and 1m in the vertical direction.

[43] Unlike the first case study, we sampled the three lithofacies fields at intervals of 20 m. Along each of the boreholes, we randomly generated electrical resistivity and seismic velocity using a Gaussian random generator with mean and standard deviation values determined according to the collocated lithofacies and the data given in Table 2. The mean values of resistivity in Table 2 were obtained from resistivity logs collected from w1250 at the LLNL site, and the mean values of seismic velocity were chosen based on published parameter ranges for unconsolidated saturated sediments [*Lankston*, 1990; *Hyndman et al.*, 1994]. We changed the complexity of the cross correlations by adjusting the standard deviations. Figure 10 shows the cross plots of the synthetic data at the five boreholes. It is evident that the nonlinearity of the cross correlations increases with the number of lithofacies types.

5.2. Approach

[44] In this case study, we followed an approach similar to the one used by *Chen et al.* [2001]. Data at each of the five boreholes were in turn to be considered as a testing set, and

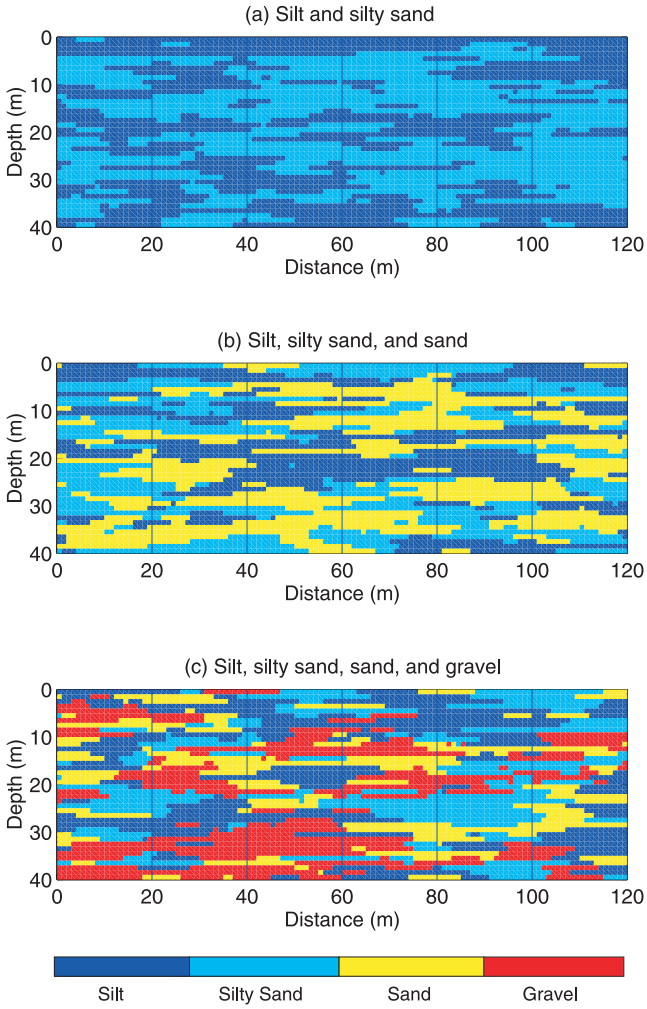

Figure 9. Synthetic lithofacies fields (a) with silt and silty sand, (b) with silt, silty sand, and sand, and (c) with silt, silty sand, sand, and gravel.

data at the corresponding other four boreholes were taken as a training set. We first used the data in the training set to train BFNN and indicator cokriging models and then used the trained models to estimate lithofacies at each testing location. The estimated values were compared to the corresponding true values, and the performance of each model was evaluated by analyzing the percentages of misclassification.

5.3. Results

[45] Table 3 summarizes the misclassifications for each data set shown in Figure 10. For the case of two types of

Table 2. Means and Standard Deviations of Geophysical Data for Each Type of Lithofacies

	Resistivity (Ωm)		Seismic Velocity, m/s	
Lithofacies	Mean	Standard Deviation	Mean	Standard Deviation
Silt	10.43	1.23	1520	130.0
Silty sand	12.89	2.10	1780	105.0
Sand	18.81	2.90	1830	105.0
Gravel	25.65	1.42	1620	105.0

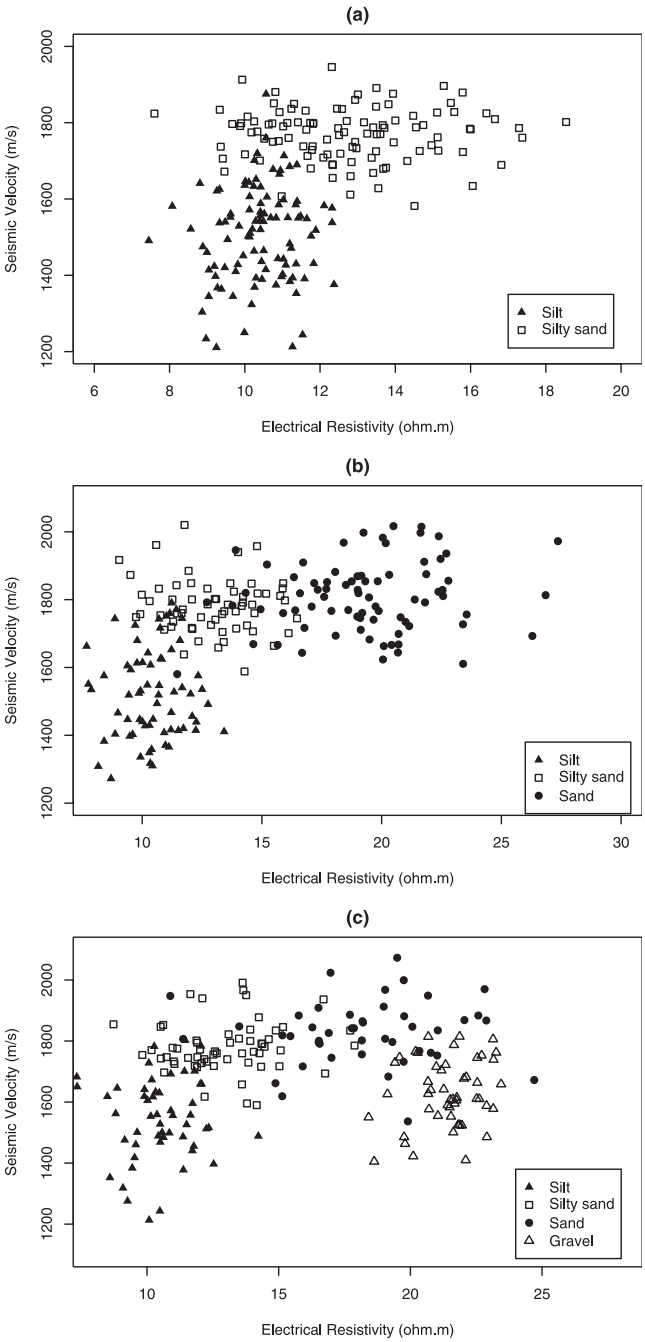

Figure 10. Cross plots of electrical resistivity versus seismic velocity (a) with two types of lithofacies, (b) with three types of lithofacies, and (c) with four types of lithofacies.

lithofacies, the nonlinearity of the cross correlation is not high, and the difference in misclassification between BFNN and indicator cokriging models is not significant. These results are consistent with the ones in the first case study. As the number of lithofacies types or the nonlinearity of the cross correlation increases, however, the differences between the two models become evident. Similar results can also be obtained if the interval between the sampled boreholes is reduced.

6. Conclusions

[46] In this study, we developed an effective Bayesian model for lithofacies estimation. The prior probability is estimated from lithofacies measurements using indicator kriging based on spatial correlation of lithofacies. The posterior probability is updated from the prior using geophysical data through the likelihood function. The effectiveness of the model in combining geophysical data comes from the use of a fuzzy neural network for the inference of likelihood functions.

[47] We used a fuzzy neural network as a nonparametric model to derive likelihood functions from cross correlations between lithofacies and geophysical attributes. These relations are typically complicated and site-specific due to the difference in measurement scales of lithofacies and geophysical data and due to uncertainty associated with acquisition and interpretation of the geophysical data. Compared to covariance-based likelihood functions, this method is more flexible and powerful in dealing with field data because it requires much less assumptions about the form of the likelihood functions.

[48] We demonstrated the Bayesian model coupled with a fuzzy neural network (BFNN) in the study using two sets of synthetic data. Results showed that BFNN is more effective than indicator kriging, indicator cokriging, and the fuzzy neural networks without using spatial correlation (FNN) in handling nonlinear correlation. Each of the alternatives can be considered as a special case of BFNN in different situations. BFNN is similar to the indicator kriging when estimating locations are close to measurements, similar to FNN when estimating locations are far from measurements, and similar to the indicator cokriging when the number of lithofacies types is less than three or the nonlinearity of cross correlation is not high.

[49] BFNN is particularly useful in cases where the nonlinearity of cross correlation between lithofacies and geo-

Table 3. Percentages of Misclassification Using Different Approaches

Testing Wells	Indicator Kriging	Indicator Cokriging	BFNN
	Case 1		
Well-1	54	5	7
Well-2	39	7	5
Well-3	39	5	2
Well-4	36	2	0
Well-5	46	12	5
Average	43	6	4
Standard deviation	7	6	4
	Case 2		
Well-1	63	20	24
Well-2	56	22	7
Well-3	39	12	10
Well-4	46	10	7
Well-5	49	29	15
Average	51	19	13
Standard deviation	9	8	7
	Case 3		
Well-1	76	37	24
Well-2	66	39	10
Well-3	68	24	22
Well-4	71	34	24
Well-5	71	41	17
Average	70	35	20
Standard deviation	4	7	6

physical attributes is high and estimating locations are within two or three integral lengths of lithofacies. BFNN is also efficient in handling multi-dimensional data sets because it uses a fuzzy neural network to infer likelihood functions; this allows for identification of complex patterns inherent in the multi-dimensional data sets, which are otherwise very difficult to obtain.

[50] BFNN has limitations. We assume that collocated geophysical data are available at any estimating location and the Markov conditions are satisfied. This is only applicable when we estimate lithofacies along geophysical tomographic profiles or when there are many borehole measurements and cross-hole geophysical data so that geophysical data can be interpolated to any estimating location. Another limitation is that although we use a nonparametric method to infer likelihood functions, the prior is still assumed Gaussian distribution and derived using kriging. It is possible to develop a nonparametric method, such as the minimum relative entropy method developed by *Woodbury and Rubin* [2000] to estimate the prior.

Appendix A: Algorithm for Parameter Identification

[51] We jointly use the least squares and the Levenberg-Marquardt methods [*McKeown*, 1980] in this study to identify parameters because $h(\mathbf{x}, \boldsymbol{\theta})$ in Equation (10) is a linear function of o_1 and o_2. The method is referred to as the hybrid model and has been demonstrated to be very efficient [*Takagi and Sugeno*, 1985; *Jang*, 1993; *Nikravesh*, 1998]. In the model, parameters o_1 and o_2 are determined using the least squares method, and other parameters, including c_{i1}, c_{i2}, σ_{i1} and σ_{i2}, $i = 1, 2$, are estimated using the Levenberg-Marquardt method. Let $\mathbf{o} = (o_1, o_2)^T$, $\boldsymbol{\alpha} = (c_{11}, c_{12}, c_{21}, c_{22}, \sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22})^T$, and *RSS* be the residual sum of squares as given below:

$$RSS(\mathbf{o}, \alpha) = \sum_{i \in A} [h(\mathbf{x}_i, \mathbf{o}, \alpha) - l(\mathbf{x}_i)]^2. \tag{A1}$$

We first estimate parameter $\mathbf{o}^{(1)}$, where the superscript denotes the iteration number, using the least squares method given the initial value $\boldsymbol{\alpha}^{(0)}$, obtained from the previous structure identification. We then estimate parameter $\boldsymbol{\alpha}^{(1)}$, using the Levenberg-Marquardt method, given the newly updated $\mathbf{o}^{(1)}$. Finally, we compute the most recently obtained residual sum of squares (*RSS*) and compare it with the previous one. The iterations are terminated when *RSS* ceases to change.

A1. Least Squares Estimation Method

[52] The least squares method used in the hybrid model is straightforward. Given parameter α, relative weights $\bar{w}_1(\mathbf{x}_i)$ and $\bar{w}_2(\mathbf{x}_i)$ for each input $(g_1(\mathbf{x}_i), g_2(\mathbf{x}_i))$, $i \in A$, can be calculated to get a matrix $\mathbf{w}$ as follows,

$$\begin{bmatrix} \bar{w}_1(\mathbf{x}_1) & \bar{w}_2(\mathbf{x}_1) \\ \bar{w}_1(\mathbf{x}_2) & \bar{w}_2(\mathbf{x}_2) \\ \vdots & \vdots \\ \bar{w}_1(\mathbf{x}_n) & \bar{w}_2(\mathbf{x}_n) \end{bmatrix}. \tag{A2}$$

Let $\mathbf{l} = (l(\mathbf{x}_1), l(\mathbf{x}_2), \cdots, l(\mathbf{x}_n))^T$. The relation between the consequence of each fuzzy rule and the output of the system is given by $\mathbf{wo} = \mathbf{l}$, and the least squares estimate of parameter $\mathbf{o}$ is given by $(\mathbf{w}^T\mathbf{w})^{-1}\mathbf{w}^T\mathbf{l}$ [*Stone*, 1995].

A2. Levenberg-Marquardt Method

[53] The Levenberg-Marquardt algorithm is the nonlinear optimization method used in this study to estimate parameter α for a given parameter $\mathbf{o}$. It is a revised Gaussian-Newton method and requires iteratively evaluating residuals $\mathbf{e} = (e_1, e_2, \cdots, e_n)^T$, where $e_i = h(\mathbf{x}_i, \mathbf{o}, \boldsymbol{\alpha}) - l(\mathbf{x}_i)$, $i \in A$, and Jacobian matrix $\mathbf{J}$ given by [*McKeown*, 1980]

$$\begin{bmatrix} \frac{\partial e_1}{\partial c_{11}} & \frac{\partial e_1}{\partial c_{12}} & \frac{\partial e_1}{\partial c_{21}} & \frac{\partial e_1}{\partial c_{22}} & \frac{\partial e_1}{\partial \sigma_{11}} & \frac{\partial e_1}{\partial \sigma_{12}} & \frac{\partial e_1}{\partial \sigma_{21}} & \frac{\partial e_1}{\partial \sigma_{22}} \\ \frac{\partial e_2}{\partial c_{11}} & \frac{\partial e_2}{\partial c_{12}} & \frac{\partial e_2}{\partial c_{21}} & \frac{\partial e_2}{\partial c_{22}} & \frac{\partial e_2}{\partial \sigma_{11}} & \frac{\partial e_2}{\partial \sigma_{12}} & \frac{\partial e_2}{\partial \sigma_{21}} & \frac{\partial e_2}{\partial \sigma_{22}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \frac{\partial e_n}{\partial c_{11}} & \frac{\partial e_n}{\partial c_{12}} & \frac{\partial e_n}{\partial c_{21}} & \frac{\partial e_n}{\partial c_{22}} & \frac{\partial e_n}{\partial \sigma_{11}} & \frac{\partial e_n}{\partial \sigma_{12}} & \frac{\partial e_n}{\partial \sigma_{21}} & \frac{\partial e_n}{\partial \sigma_{22}} \end{bmatrix}, \tag{A3}$$

where

$$\frac{\partial e_i}{\partial c_{ks}} = 2\bar{w}_k(\mathbf{x}_i)[o_k - h(\mathbf{x}_i, \mathbf{o}, \alpha)] \cdot \frac{[g_s(\mathbf{x}_i) - c_{ks}]}{\sigma_{ks}^2}, \quad k, s = 1, 2,$$

$$\frac{\partial e_i}{\partial \sigma_{ks}} = 2\bar{w}_k(\mathbf{x}_i)[o_k - h(\mathbf{x}_i, \mathbf{o}, \alpha)] \cdot \frac{[g_s(\mathbf{x}_i) - c_{ks}]^2}{\sigma_{ks}^3}, \quad k, s = 1, 2, \tag{A4}$$

which are obtained using the chain rule of differentiation and the membership functions given in Equation 8. The detailed algorithm is as follows.

[54] 1. Compute residual $\mathbf{e}$ and Jacobian matrix $\mathbf{J}$ for given parameters $\boldsymbol{\alpha}^{(0)}$ and set $t = 0$.

[55] 2. Check gradient $\mathbf{J}^T\mathbf{e}$. If $|\mathbf{J}^T\mathbf{e}| < \varepsilon$, stop; otherwise, go to next step.

[56] 3. For a given positive value γ, compute $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)} - (\mathbf{J}^T\mathbf{J} + \gamma I)^{-1}\mathbf{J}^T\mathbf{e}$, where I is an identity matrix, and check whether $\boldsymbol{RSS}(\mathrm{o}^{(t)}, \boldsymbol{\alpha}^{(t)}) < \boldsymbol{RSS}(\mathrm{o}^{(t)}, \boldsymbol{\alpha}^{(t-1)})$. If it does, go to next step; otherwise increase the γ value and repeat step 3 until the *RSS* is reduced.

[57] 4. Check $|\boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}^{(t-1)}| < \epsilon$. If it does, stop; otherwise, set t = t + 1 and repeat step 1.

Appendix B: Indicator Cokriging

[58] Let $\boldsymbol{g_1(x)}$ be the gamma ray shaliness and $\boldsymbol{g_2(x)}$ the normalized electrical resistivity at location $\mathbf{x}$. The conditional probability of the $\boldsymbol{k}$th lithofacies occurring at location $\mathbf{x}$ is given by

$$\begin{aligned} \boldsymbol{P}(\boldsymbol{Z}(\mathbf{x}) = \boldsymbol{k} | \boldsymbol{g}_1(\mathbf{x}), \boldsymbol{g}_2(\mathbf{x})) = \boldsymbol{p_k} &+ \sum_{\boldsymbol{i} \in \boldsymbol{A}} \lambda_{\boldsymbol{i}}(\mathbf{x})(\boldsymbol{I_k}(\mathbf{x}_i) - \boldsymbol{p_k}) \\ &+ \boldsymbol{s}_1(\mathbf{x})(\boldsymbol{g_1}(\mathbf{x}) - \boldsymbol{m_1}) \\ &+ \boldsymbol{s}_2(\mathbf{x})(\boldsymbol{g}_2(\mathbf{x}) - \boldsymbol{m}_2), \end{aligned} \tag{B1}$$

and

$$\sum_{i\in A}\lambda_i(\mathbf{x})\rho_I(\mathbf{x}_i,\mathbf{x}_j)+s_1(\mathbf{x})\rho_{Ig1}(\mathbf{x},\mathbf{x}_j)+s_2(\mathbf{x})\rho_{Ig2}(\mathbf{x},\mathbf{x}_j)$$
$$=\rho_I(\mathbf{x},\mathbf{x}_j),\qquad j\in A,$$
$$\sum_{i\in A}\lambda_i(\mathbf{x})\rho_{Ig1}(\mathbf{x},\mathbf{x}_i)+s_1(\mathbf{x})\rho_{g1}+s_2(\mathbf{x})\rho_{g1g2}=\rho_{Ig1},$$
$$\sum_{i\in A}\lambda_i(\mathbf{x})\rho_{Ig2}(\mathbf{x},\mathbf{x}_i)+s_1(\mathbf{x})\rho_{g1g2}+s_2(x)\rho_{g2}=\rho_{Ig2},\qquad \text{(B2)}$$

where m_1 and m_2 are the means of the gamma ray shaliness and the normalized resistivity, and s_1 and s_2 are the cokriging coefficients of the gamma ray shaliness and the normalized resistivity. ρ_I, ρ_{g1} and ρ_{g2} are the correlation coefficients of the indicator variable, the gamma ray shaliness, and the electrical resistivity, respectively. ρ_{Ig1}, ρ_{Ig2}, and ρ_{g1g2} are the cross correlation coefficients between the indicator variable and the geophysical data $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$.

[59] **Acknowledgments.** Funding for this study was provided by the National Science Foundation (NSF) through grant EAR 9628306. We thank R. Bainer and R. Blake from Lawrence Livemore National Laboratory, D. G. Hill and S. Ezzedine from Weiss Associates, and S. Hubbard and M. Kowalsky from the Lawrence Berkeley National Laboratory. We also thank two anonymous reviewers for providing insightful and useful suggestions that helped to improve the manuscript.

References

Aarts, E., and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley, New York, 1989.

Almeida, A. S., and A. G. Journel, Joint simulation of multiple variables with a Markov-type coregionalization model, *Math. Geol.*, *26*, 565–587, 1994.

Banfield, J. D., and A. E. Raftery, Model-based Gaussian and non-Gaussian clustering, *Biometrics*, *49*, 803–821, 1993.

Bernardo, J. M., and A. F. Smith, *Bayesian Theory*, John Wiley, New York, 1994.

Bezdek, J., *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.

Bosch, M., Lithologic tomography: From plural geophysical data to lithology estimation, *J. Geophys. Res.*, *104*, 749–766, 1999.

Bosch, M., A. Guillen, and P. Ledru, Lithologic tomography: An application to geophysical data from the Cadomian Belt of Northern Brittany, France, *Tectonophysics*, *31*, 197–227, 2001.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley-Longman, Reading, Mass., 1973.

Chang, H. C., H. C. Chen, and J. H. Fang, Lithology determination from well logs with fuzzy associative memory neural networks, *IEEE Trans. Geosci. Remote Sens.*, *35*, 773–780, 1997.

Chen, J., S. Hubbard, and Y. Rubin, Estimating the hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model, *Water Resour. Res.*, *37*, 1603–1613, 2001.

Copty, N., and Y. Rubin, A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, *31*, 1673–1686, 1995.

Dagan, G., *Flow and Transport in Porous Formations*, Springer-Verlag, New York, 1989.

Deutsch, C. V., and P. W. Cockerham, Practical considerations in the application of simulated annealing to stochastic simulation, *Math. Geol.*, *26*, 67–82, 1994.

Deutsch, C., and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York, 1998.

Doveton, J. H., *Log Analysis of Subsurface Geology Concepts and Computer Methods*, John Wiley, New York, 1986.

Ezzedine, S., Y. Rubin, and J. Chen, Hydrological-geophysical Bayesian method for subsurface site charaterization: Theory and application to LLNL Superfund Site, *Water Resour. Res.*, *35*, 2671–2683, 1999.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, 1998.

Hartigan, J. A., and M. A. Wong, A K-means clustering algorithm, *Appl. Stat.*, *28*, 100–108, 1979.

Hirota, K., and W. Pedrycz, Fuzzy computing for data mining, *Proc. IEEE*, *87*, 1575–1599, 1999.

Hubbard, S., and Y. Rubin, Hydrogeological parameter estimation using geological data: A review of selected techniques, *J. Contam. Hydrol.*, *45*, 3–34, 2000.

Hyndman, D. W., J. M. Harris, and S. M. Gorelick, Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, *30*, 1965–1977, 1994.

Jang, J. S., ANFIS: Adaptive network based fuzzy inference system, *IEEE Trans. Syst. Man Cybernetics*, *23*, 665–685, 1993.

Keys, W. S., *A Practical Guide to Borehole Geophysics in Environmental Investigations*, A. F. Lewis, New York, 1997.

Kitanidis, P. K., Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, *22*, 499–507, 1986.

Kitanidis, P. K., *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge Univ. Press, New York, 1997.

Kitanidis, P. K., and E. G. Vomvoris, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations, *Water Resour. Res.*, *19*, 677–690, 1983.

Koltermann, C. E., and S. M. Gorelick, Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, *32*, 2617–2658, 1996.

Kowalsky, M., P. Dietrich, G. Teutsch, and Y. Rubin, Forward modeling of ground-penetrating radar data using digitized outcrop images and multiple scenarios of water saturation, *Water Resour. Res.*, *37*, 1615–1625, 2001.

Lankston, R. W., High-resolution refraction seismic data acquisition and interpretation, in *Geotechnical and Environmental Geophysics*, edited by S. H. Ward, Soc. of Explor. Geophys., Tulsa, Okla., 1990.

Lortzer, G. J. M., and A. J. Berkhout, An integrated approach to lithologic inversion, part I, Theory, *Geophysics*, *57*, 233–244, 1992.

McKeown, J., Nonlinear least-squares problems, in *Nonlinear Optimization: Theory and Algorithm*, edited by L. Dixon, E. Spedicato, and G. Szego, pp. 91–105, Birkhäuser Boston, Cambridge, Mass., 1980.

Nikravesh, M., Neural network knowledge-based modeling of rock properties based on well log databases, paper presented at BISC-SIG-ES Seminar, Univ. of Calif., Berkeley, 1998.

Rogers, S. J., J. Fang, C. Karr, and D. Stanley, Determination of lithology from well logs using a neural network, *AAPG Bull.*, *76*, 731–739, 1992.

Rojas, R., *Neural Networks: A Systematic Introduction*, Springer-Verlag, New York, 1996.

Rosenbaum, M., L. Rosen, and G. Gustafson, Probabilistic models for estimating lithology, *Eng. Geol.*, *47*, 43–55, 1997.

Rubin, Y., *Applied Stochastic Hydrology*, Cambridge Univ. Press, New York, 2003.

Rubin, Y., and G. Dagan, Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 1, Constant head boundary, *Water Resour. Res.*, *24*, 1689–1697, 1988.

Rubin, Y., and G. Dagan, Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 2, Impervious boundary, *Water Resour. Res.*, *25*, 707–712, 1989.

Rubin, Y., G. Mavko, and J. Harris, Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, *Water Resour. Res.*, *28*, 1809–1816, 1992.

Stone, C. J., *A Course in Probability and Statistics*, Duxbury, New York, 1995.

Takagi, T., and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Syst. Man Cybernetics*, *15*, 116–132, 1985.

Woodbury, A. D., and Y. Rubin, A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod Experimental Site, *Water Resour. Res.*, *36*, 159–171, 2000.

Xie, X., and G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. and Mach. Intell.*, *13*, 841–847, 1991.

Zadeh, L. A., Fuzzy sets, *Inf. Control*, *8*, 338–353, 1965.

Zadeh, L. A., Soft computing and fuzzy logic, *IEEE Software*, *11*, 48–56, 1994.

J. Chen and Y. Rubin, Department of Civil and Environmental Engineering, University of California, Berkeley, 440 Davis Hall, Berkeley, CA 94720-1710, USA. (jchen@lbl.gov)